In transit data processing for extreme scale computing

Scott A. Klasky

+

H. Abbasi[2], Q. Liu[2], J. Logan[1], M. Parashar[6], N. Podhorszki[1], K. Schwan[4], A. Shoshani[3], M. Wolf[4], S. Ahern[1], I.Altintas[9], W.Bethel[3], L. Chacon[1], CS Chang[10], J. Chen[5], H. Childs[3], J. Cummings[13], C. Docan[6], G. Eisenhauer[4], S. Ethier[10], R. Grout[7], R. Kendall[1], J. Kim[3], T. Kordenbrock[5], Z. Lin[10], J. Lofstead[5], X. Ma[15], K. Moreland[5], R. Oldfield[5], V. Pascucci[12], N. Samatova[15], W. Schroeder[8], R. Tchoua[1], Y. Tian[14], R. Vatsavai[1], M. Vouk[15], K. Wu[3], W. Yu[14], F. Zhang[6], F. Zheng[4] + many many many more!

[1]ORNL, [2] U.T. Knoxville, [3]LBNL, [4]Georgia Tech, [5]Sandia Labs, [6] Rutgers, [7]NREL, [8]Kitware, [9]UCSD, [10]PPPL, [11]UC Irvine, [12]U. Utah, [13] Caltech, [14]Auburn University, [15]NCSU

Managed by UT-Battelle
for the Department of Energy

OLCF
Oak Ridge Leadership Computing Facility

Main points I want to make

- Creating a computational laboratory is really difficult
  - Problems are really hard, Multi-scale, Multi-physics
- Complexities come from
  - Systems
  - Applications
  - Rapidly changing requirements
  - Evolving target platforms
  - Diverse Team
- And we never heard this before….

It's all new this time around?

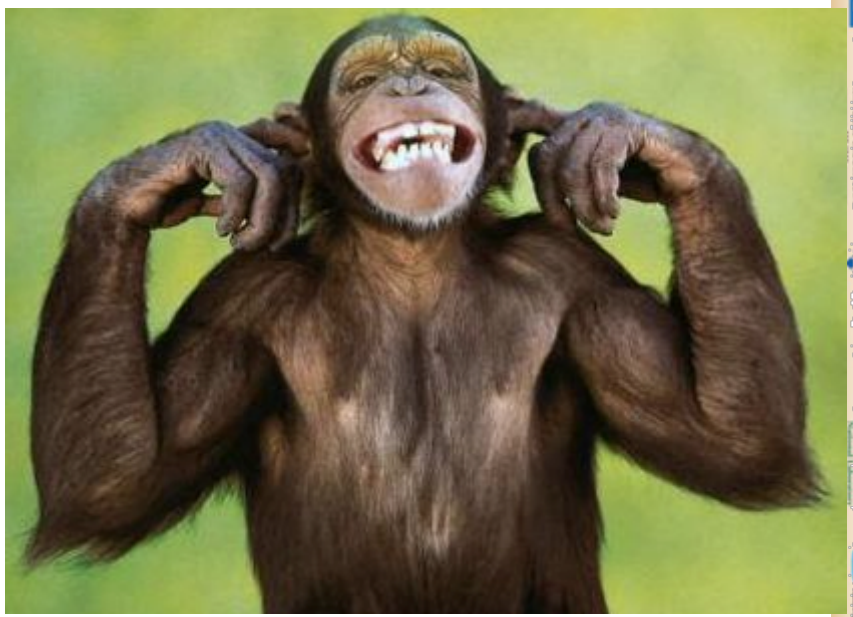
What is so difficult?

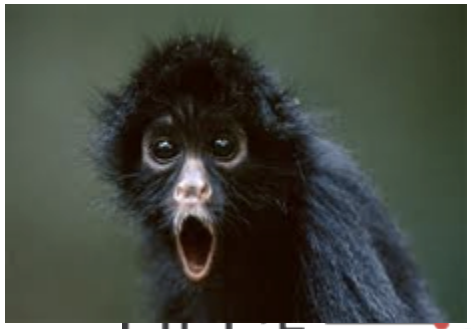
- The problem has at least multiple scales
  - -natural scale near each hole ~m
  - -ringdown natural scale ~20 $(m_1 + m_2)$
  - -inspiral wavelength ~ hundreds x $(m_1 + m_2)$

Need Approach to Handle Different Scales ! ! ! !

Some people think I/O is only about reads/writes

- But why do we want to read/ write?
- Is checkpoint-restart it?
- Some people say that you can re-run your code, and place analytics inside the code.
  - $150K/day to run a code, what will the cost be for the exascale?
- Some people say that they can just program I/O + analytics on their own.





Some people think we just need to collaborate

S3D +  Boxlib +  ROSE +  ADIOS +  DSL 

= 

Or

My beliefs for exascale

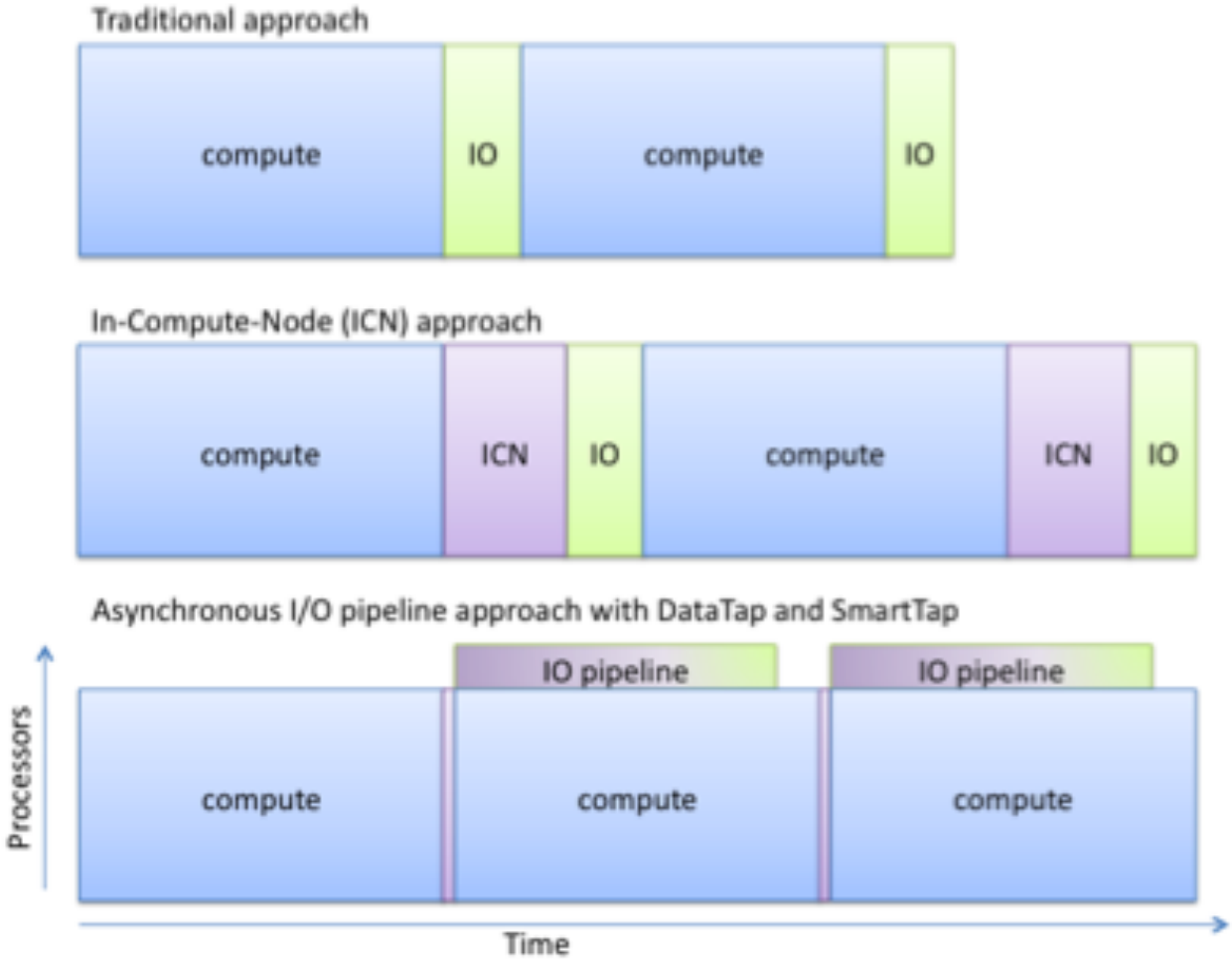
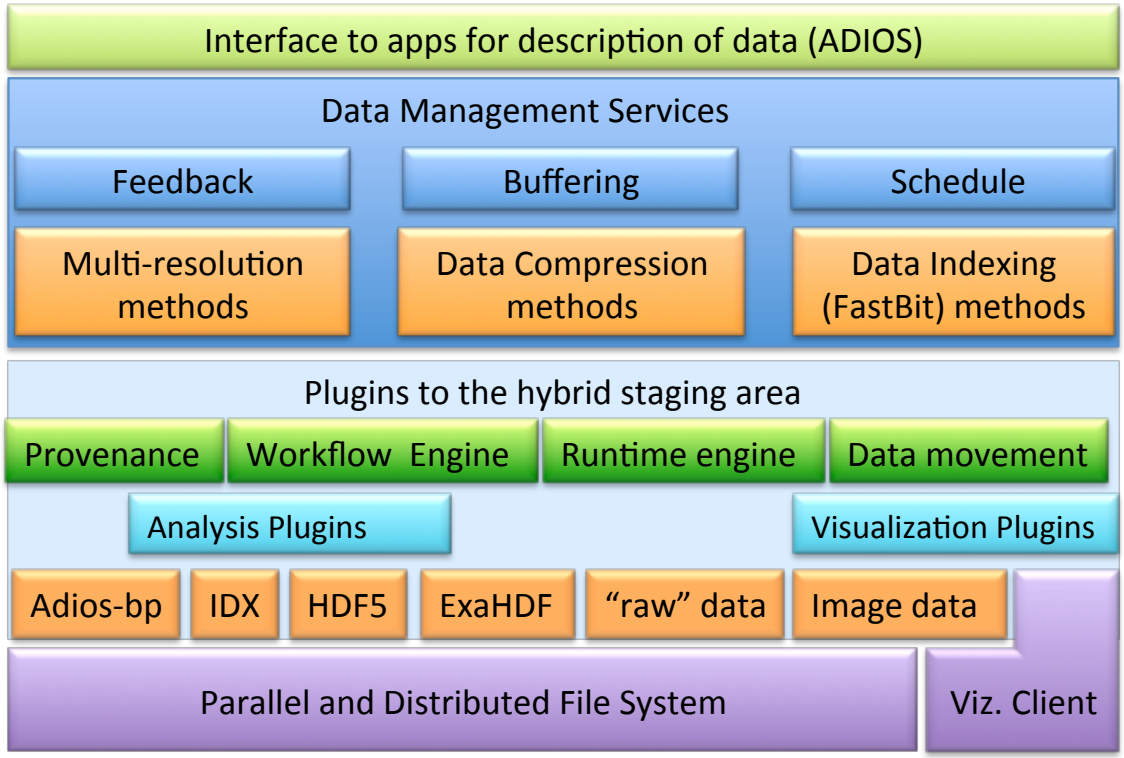


Fig. 5. Different placements of I/O activities in time and processor space

Use principles from the Enterprise

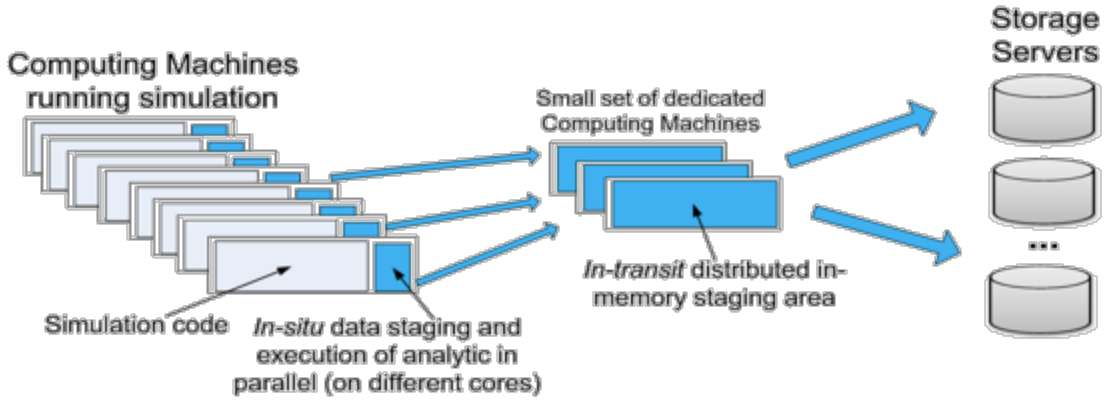
- SOA --- Yes that means Service Oriented Architecture
- But…. For HPC/ I mean Exascale.
- Our approach:



But keep an open mind to new ideas

My belief for exascale (from Exact)

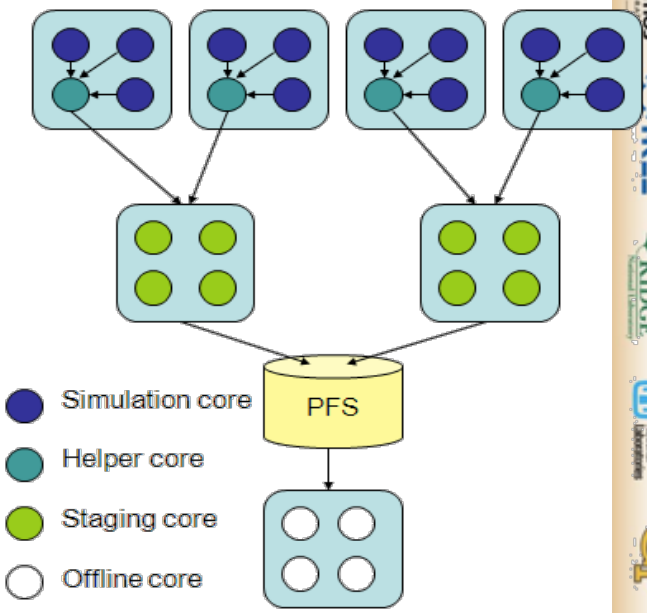
- ADIOS is being implemented as a Service Oriented Architecture
  - Create an environment where applications can abstract data movement/IO as read/writes/queries



- Exploit complex-memory hierarchies with ***Hybrid Data Staging*** to:
  - Decrease the gap between CPU and IO speeds
  - Dynamically deploy and execute data analytical or pre-processing operations either in-situ or in-transit
  - Improved IO write performance

Co-design the analytics pipeline

- **Optimization:**
- **Where/How should analytics be run?**
  - Inline with simulation?
  - Separate cores? Helper cores?
  - Separate staging nodes?
  - Offline?



- **Guided by what metrics?**
  - *Performance*: **Total Execution Time** of both simulation and analytics
  - *Cost*: **CPU hours** charged for simulation and analytics
  - *Time to Data*: **Delay** between data generation and analytics results
  - *Data movement* ~ **Power** consumption